CrossMark

Phonology **34** (2017) 243–268. © Cambridge University Press 2017
doi:10.1017/S0952675717000136

*Constraint conjunction in weighted probabilistic grammar**

Stephanie S. Shih
University of California, Merced

This paper examines a key difference between constraint conjunction and constraint weight additivity, arguing that the two do not have the same empirical coverage. In particular, constraint conjunction in weighted probabilistic grammar allows for superadditive constraint interaction, where the effect of violating two constraints goes beyond the additive combination of the two constraints' weights alone. A case study from parasitic tone harmony in Dioula d'Odienné demonstrates superadditive local and long-distance segmental feature similarities that increase the likelihood of tone harmony. Superadditivity in Dioula d'Odienné is formally captured in Maximum Entropy Harmonic Grammar by weighted constraint conjunction. Counter to previous approaches that supplant constraint conjunction with weight additivity in Harmonic Grammar, information-theoretic model comparison reveals that weighted constraint conjunction improves the grammar's explanatory power when modelling quantitative natural language patterns.

1 Introduction

PHONOLOGICAL GANGING is a type of cumulativity in constraint interaction wherein one strong constraint $\mathbb{C}_1$ can be overtaken by two weaker constraints $\mathbb{C}_2$ and $\mathbb{C}_3$ together, but not by $\mathbb{C}_2$ or $\mathbb{C}_3$ independently (e.g. Jäger & Rosenbach 2006). In traditional, strict-ranking Optimality Theory (OT; Prince & Smolensky 1993), phonological ganging is achieved via the (local) conjunction of two constraints – e.g. $\mathbb{C}_2\&\mathbb{C}_3$ – defined in (1) (e.g. Smolensky 1993, 2006, Ito & Mester 1998, 2003, Baković 2000; cf. Crowhurst & Hewitt 1997). Conjunctions of constraints usually have locality-domain restrictions (e.g. Łubowicz 2005), and are assumed to rank above their simplex counterparts: $\mathbb{C}_2\&\mathbb{C}_3 \gg \mathbb{C}_2, \mathbb{C}_3$ (e.g. Baković 2000: 27ff).

(1) $\mathbb{C}_2\&_D\mathbb{C}_3$
Assign a violation if both $\mathbb{C}_2$ and $\mathbb{C}_3$ are violated (in local domain D).

* E-mail: STEPHSUS@GMAIL.COM.
Acknowledgements to Kie Zuraw, Eric Baković, Christopher Green, Lauren Hall-Lew, Gunnar Hansson, Sharon Inkelas, Laura McPherson, Joe Pater, Brian Smith, Anne-Michelle Tessier, Colin Wilson and audiences at WCCFL 33, GLOW 38 and UCs Merced and Berkeley for discussion on various portions of this work. Thanks also to Jeff Heinz and Bill Idsardi, and two anonymous reviewers for detailed comments. Errors are inevitable in stochastic systems, and the author accepts blame for any that remain herein.

https://doi.org/10.1017/S0952675717000136 Published online by Cambridge University Press

The pair of tableaux in (2) illustrates ganging using constraint conjunction. In (2a), the winning candidate violates only one of the lower-ranked constraints, $\mathbb{C}_2$ or $\mathbb{C}_3$, and the losing candidate violates the higher-ranked constraint, $\mathbb{C}_1$. But if both $\mathbb{C}_2$ and $\mathbb{C}_3$ are violated, as in candidate (2b.i), then the higher-ranked conjunction of $\mathbb{C}_2 \& \mathbb{C}_3$ is also violated. In this way, the combined ganging effect of violating both $\mathbb{C}_2$ and $\mathbb{C}_3$ overtakes $\mathbb{C}_1$.

(2) a.

	$\mathbb{C}_2 \& \mathbb{C}_3$	$\mathbb{C}_1$	$\mathbb{C}_2$	$\mathbb{C}_3$
☞ i. Winner			1	
ii. Loser		1		

b.

	$\mathbb{C}_2 \& \mathbb{C}_3$	$\mathbb{C}_1$	$\mathbb{C}_2$	$\mathbb{C}_3$
i. Loser	1		1	1
☞ ii. Winner		1		

In contrast to strict-ranking OT, weighted constraint approaches in Harmonic Grammar (HG; e.g. Legendre *et al*. 1990, Smolensky & Legendre 2006) can achieve ganging via cumulative constraint addition, leading many to posit that constraint conjunction may not be a necessary mechanism in HG (e.g. Farris-Trimble 2008, Potts *et al*. 2010, Pater 2016). The HG tableaux in (3) illustrate the same ganging effect as shown in (2). In (3a), the winning candidate violates only one of the lower-weighted constraints, $\mathbb{C}_2$ or $\mathbb{C}_3$, incurring a lower harmony score ($\mathcal{H}$) than the losing candidate, which violates the higher-weighted constraint $\mathbb{C}_1$. Ganging occurs when both $\mathbb{C}_2$ and $\mathbb{C}_3$ are violated, as in (3b): the sum of the weighted violations of the simplex constraints $\mathbb{C}_2$ and $\mathbb{C}_3$ is additively greater than a single violation of $\mathbb{C}_1$. So the candidate that violates both lower-weighted constraints, (i), loses to the one that only violates the higher-weighted constraint, (ii).

(3) a.

	$\mathbb{C}_1$ 3	$\mathbb{C}_2$ 2	$\mathbb{C}_3$ 2	$\mathcal{H}$
☞ i. Winner		1		2
ii. Loser	1			3

b.

	$\mathbb{C}_1$ 3	$\mathbb{C}_2$ 2	$\mathbb{C}_3$ 2	$\mathcal{H}$
i. Loser		1	1	4
☞ ii. Winner	1			3

This paper argues that additive constraint cumulativity in HG does not entirely supplant constraint conjunction, and that constraint conjunction is needed in HG to capture quantitative phonological patterns. At least two major points of difference between constraint conjunction and additive cumulativity have been previously noted. The first is structural, in that additive cumulativity in HG models locality conditions differently than constraint conjunction does (Pater 2016; on locality conditions in conjunction, see e.g. Baković 2000, Łubowicz 2005). The second point of difference – and the focus of this paper – is that conjunction in HG has the ability to capture superadditivity effects in quantitative phonological patterns (suggested by Albright 2009, and also implemented by Green & Davis 2014). Superadditivity occurs when the cumulative effect of two constraints is greater than that of the two simplex constraints. It can be captured in HG as WEIGHTED CONSTRAINT CONJUNCTION, in which a

https://doi.org/10.1017/S0952675717000136 Published online by Cambridge University Press

conjoined constraint, $\mathbb{C}_2\&\mathbb{C}_3$, receives an independent weight, above and beyond the additive effects of the simplex constraints $\mathbb{C}_2$ and $\mathbb{C}_3$. This is illustrated in (4), where the conjunction $\mathbb{C}_2\&\mathbb{C}_3$ is assigned a weight which is independent of the weights of $\mathbb{C}_2$ and $\mathbb{C}_3$. The losing candidate, which violates the conjunction $\mathbb{C}_2\&\mathbb{C}_3$ and the simplex constraints $\mathbb{C}_2$ and $\mathbb{C}_3$, receives a harmony score of 6. Compared to the HG tableau in (3b), which does not have weighted constraint conjunction, the winning candidate in (4) wins by a larger harmony-score difference ($\Delta\mathcal{H} = 3$) than the winning candidate in (3b) does ($\Delta\mathcal{H} = 1$).

(4)

	$\mathbb{C}_2\&\mathbb{C}_3$ 2	$\mathbb{C}_1$ 3	$\mathbb{C}_2$ 2	$\mathbb{C}_3$ 2	$\mathcal{H}$
a. Loser	1		1	1	6
☞ b. Winner		1			3

Being able to capture superadditive effects using weighted constraint conjunction in HG has the advantage of providing more explanatory power in modelling patterns found in variable natural language data. This paper presents evidence from a case from the tone system of Dioula d'Odienné (Mande, Ivory Coast; henceforth Dioula) that illustrates such a need for superadditivity. An information-theoretic methodology of model selection and comparison is introduced to assess the contribution of weighted constraint conjunction to the grammar, measuring the trade-off between the loss of restrictiveness that comes with adding conjoined constraints to CON and the gain of predictive accuracy over the observed data (see Wilson & Obdeyn 2009 for a similar approach). The main message here is that we suffer a potentially significant loss of information and explanatory power if grammars are *a priori* restricted from having constraint conjunction. Instead, the necessity and viability of conjunction must be quantitatively assessed against noisy natural language data.

The paper is organised as follows. §2 introduces the implementation of weighted constraint conjunction used here, with toy grammars to demonstrate the differences between HG approaches with and without conjunction. §3 presents data from a tone-alternation phenomenon in Dioula d'Odienné nouns that illustrates the need for superadditivity. In §4, grammars for Dioula with and without constraint conjunction are compared, using information-theoretic model selection. §5 discusses how the weight of evidence test can also identify cases in which superadditivity is not necessary. §6 concludes.

2 Modelling cumulative ganging

To capture quantitative phonological patterns, I use Maximum Entropy Harmonic Grammar (henceforth MaxEnt; Goldwater & Johnson 2003, Wilson 2006, Jäger 2007, Hayes & Wilson 2008, among others). Like non-stochastic HG, MaxEnt uses weighted constraints; however, instead

https://doi.org/10.1017/S0952675717000136 Published online by Cambridge University Press

of winner and loser candidates in OT and non-stochastic HG, a MaxEnt grammar produces a probability distribution over the candidate set based on constraint weights. As such, MaxEnt grammars are able to model both variable and (near-)categorical distributions. From candidates' harmony scores, which are based on multiplying constraint violations by respective constraint weights, (5a), the probability of a given candidate can be calculated over all other possible candidates for the given input, (5b).

(5) a. $\mathcal{H}(y|x) = \sum_{i=1}^{n} w_i\, \mathbb{C}_i(y,x) = \boldsymbol{w_i \times \mathbb{C}_i}$

where y is the candidate for input x
w_i is the weight of constraint $\mathbb{C}_i$
$\mathbb{C}_i(y,x)$ is the number of violations of $\mathbb{C}_i$ that (y,x) incurs
n is the maximum value of i in constraints $(\mathbb{C}_i \ldots \mathbb{C}_n)$

b. $P(y|x) = \exp(-\mathcal{H}(y|x)) / Z(x)$

where $Z(x) = \sum_{y \in \mathcal{Y}(x)} \exp(-\mathcal{H}(y|x))$
$\mathcal{Y}(x)$ is the set of candidate output forms given input x

The MaxEnt results reported herein were fitted using the MaxEnt Grammar Tool (Hayes *et al*. 2009). MaxEnt grammars are part of a general class of statistical models, also known as log-linear or multinomial logistic regression models. One primary difference between MaxEnt HG and regression models is the requirement of non-negative MaxEnt constraint weights imposed by optimality-theoretic assumptions. Consequences for conjunction when this requirement is relaxed in logistic regression models are discussed in online supplementary materials.[1]

Weighted constraint conjunction for superadditive ganging effects is akin to a standard feature in regression models: interaction terms. In regression modelling, interactions occur when the effect of two constraints on a third is not merely additive – that is, when the combination of two constraints has an effect over and above the mere addition of those two constraints. Such interactions are commonplace in linguistic systems and analyses, and have been demonstrated in numerous phonetic, sociolinguistic, psycholinguistic and corpus studies that model quantitative linguistic patterns computationally (e.g. Jäger & Rosenbach 2006).

Interactions are implemented in regression models as the product of two constraints, $\mathbb{C}_1 \times \mathbb{C}_2$, which can receive its own weight w, independent of the weights assigned to $\mathbb{C}_1$ and $\mathbb{C}_2$ separately, as shown in (6).[2]

(6) $w_1\mathbb{C}_1 + w_2\mathbb{C}_2 + w_3(\mathbb{C}_1 \times \mathbb{C}_2) + \ldots + w_n\mathbb{C}_n = \boldsymbol{w_i \times \mathbb{C}_i} = \mathcal{H}(x,i)$

When the possible violations of constraints are limited to {0, 1}, then the multiplicative interaction term – i.e. $\mathbb{C}_1 \times \mathbb{C}_2$ – is equivalent to previously

[1] Available at https://doi.org/10.1017/S0952675717000136.

[2] The definition in (5a) has to be adjusted such that w_i = weight of constraint $\mathbb{C}_i$ or a conjunction of constraints.

https://doi.org/10.1017/S0952675717000136 Published online by Cambridge University Press

implemented forms of constraint conjunction (i.e. $\mathbb{C}_1 \& \mathbb{C}_2$), leaving locality conditions aside. But when more than one violation is possible (or, as in OT, the number of violations is positively unbounded: [0, +∞)), then $\mathbb{C}_1 \times \mathbb{C}_2$ and $\mathbb{C}_1 \& \mathbb{C}_2$ are no longer equivalent. In the interaction implementation, $\mathbb{C}_1 \times \mathbb{C}_2$, violations are multiplied: if $\mathbb{C}_1$ receives two violations and $\mathbb{C}_2$ receives one violation, then $\mathbb{C}_1 \times \mathbb{C}_2$ has two violations. In contrast, on standard assumptions for constraint conjunction, the conjunction $\mathbb{C}_1 \& \mathbb{C}_2$ is assigned only one violation, regardless of the number of violations for $\mathbb{C}_1$ and $\mathbb{C}_2$ independently: even if $\mathbb{C}_1$ has two violations and $\mathbb{C}_2$ one, $\mathbb{C}_1 \& \mathbb{C}_2$ still only incurs a single violation. To remain consistent with statistical regression analyses, weighted constraint conjunctions are implemented in this paper as the product of two constraints. Exploring the full consequences of this multiplicative implementation of weighted constraint conjunction, however, is left for future work.

The utility of weighted constraint conjunctions to model interactions has already been demonstrated – though sparingly – in the phonological literature couched in MaxEnt and HG approaches (see Albright 2009 for a first proposal for capturing superadditivity). For example, Hayes *et al*. (2012) demonstrate interactions between positional and rhythmic constraints in metrical verse, Pater & Moreton (2012) use weighted interactions in HG for feature co-occurrences and Green & Davis (2014) implement weighted conjunction to restrict the phonotactics of complex syllable margins.

2.1 Toy grammar illustration

While additive cumulativity of the weights of simplex constraints in HG captures standard ganging phenomena well, superadditive phenomena that exhibit independent effects that are not merely additive reveals the shortcoming of additive cumulativity. The toy language in (7) illustrates this case; HG without conjunction makes inaccurate predictions, but HG with conjunction makes accurate predictions, as shown later in this section. Weights for grammars in this section were obtained using hand-weighting and the MaxEnt Grammar Tool (Hayes *et al*. 2009).

(7)

		frequency		$\mathbb{C}_1$ 0.84	$\mathbb{C}_2$ 0.84	$\mathbb{C}_3$ 0.84	$\mathcal{H}$	predicted
a. /input 1/	☞ i.	60%	Winner		1		0.84	50%
	ii.	40%	Loser	1			0.84	50%
b. /input 2/	☞ i.	60%	Winner			1	0.84	50%
	ii.	40%	Loser	1			0.84	50%
c. /input 3/	i.	10%	Loser		1	1	1.68	30%
	☞ ii.	90%	Winner	1			0.84	70%
d. /input 4/	☞ i.	70%	Winner				0	70%
	ii.	30%	Loser	1			0.84	30%

https://doi.org/10.1017/S0952675717000136 Published online by Cambridge University Press

In this toy language, candidates that violate only $\mathbb{C}_2$ or $\mathbb{C}_3$ are observed in (a, b) to be more frequent than candidates that violate $\mathbb{C}_1$, occurring at a 60:40 ratio. A candidate that violates both $\mathbb{C}_2$ and $\mathbb{C}_3$ is much less frequent than a candidate that violates only $\mathbb{C}_1$, occurring at a 10:90 ratio, as in (c). Finally, as shown in (d), a candidate that does not violate any constraint is better than one that violates $\mathbb{C}_1$, occurring at a 70:30 ratio.

Given this input pattern, the constraints are weighted evenly, because the grammar is attempting to reconcile two competing forces. First, the individual weights of $\mathbb{C}_2$ and $\mathbb{C}_3$ must be low enough to allow candidates that violate $\mathbb{C}_2$ and $\mathbb{C}_3$ individually to win in (7a, b). But at the same time, the sum of $\mathbb{C}_2$ and $\mathbb{C}_3$ must be high enough to overtake $\mathbb{C}_1$ by a large enough margin to achieve the extreme 10:90 gang effect in (c). The weight of $\mathbb{C}_1$ is kept in check by the pair in (d).[3] Under this schema, regular additive ganging cannot capture the extreme effect shown in (c), where the candidate violating both $\mathbb{C}_2$ and $\mathbb{C}_3$ loses by a large margin. The predictions of the grammar, given in the rightmost column, are wrong, because the grammar does not have enough degrees of freedom to capture the independent slope variances of the gang effect *vs.* the simplex effect.

Decreasing the weights of $\mathbb{C}_2$ and $\mathbb{C}_3$ to try to accurately map the surface frequencies for inputs 1 and 2 in the toy language results in a greater inaccuracy in predicting the extreme gang effect. This adjustment is shown in (8).

(8)

		frequency		$\mathbb{C}_1$ 0.84	$\mathbb{C}_2$ 0.44	$\mathbb{C}_3$ 0.44	$\mathcal{H}$	predicted
a. /input 1/	☞ i.	60%	Winner		1		0.44	60%
	ii.	40%	Loser	1			0.84	40%
b. /input 2/	☞ i.	60%	Winner			1	0.44	60%
	ii.	40%	Loser	1			0.84	40%
c. /input 3/	i.	10%	Loser		1	1	0.88	49%
	☞ ii.	90%	Winner	1			0.84	51%
d. /input 4/	☞ i.	70%	Winner				0	70%
	ii.	30%	Loser	1			0.84	30%

The decrease in the weights of $\mathbb{C}_2$ and $\mathbb{C}_3$ now allows for an accurate prediction in (a, b), where the winning candidates only violate $\mathbb{C}_2$ or $\mathbb{C}_3$ independently. However, the additive effect of $\mathbb{C}_2$ and $\mathbb{C}_3$ is also now insufficiently large to capture the gang effect in (c). The grammar, which relies on standard additive cumulativity, mispredicts a ratio of 49:51 between the loser and winner, whereas the observed surface pattern is 10:90. The grammar does achieve a gang effect, such that the sum of $\mathbb{C}_2$ and $\mathbb{C}_3$ is greater than $\mathbb{C}_1$, but crucially, the grammar cannot achieve *enough* of a gang effect.

[3] Thanks to Kie Zuraw for this observation.

https://doi.org/10.1017/S0952675717000136 Published online by Cambridge University Press

Weighted constraint conjunction allows the grammar to specify a separate, independent weight, above and beyond the additive effect of $\mathbb{C}_2$ and $\mathbb{C}_3$. A grammar with weighted constraint conjunction is given in (9).

(9)

		frequency		$\mathbb{C}_2$ & $\mathbb{C}_3$ 2.16	$\mathbb{C}_1$ 0.84	$\mathbb{C}_2$ 0.44	$\mathbb{C}_3$ 0.44	$\mathcal{H}$	predicted
a. /input 1/	☞ i.	60%	Winner			1		0.44	60%
	ii.	40%	Loser		1			0.84	40%
b. /input 2/	☞ i.	60%	Winner				1	0.44	60%
	ii.	40%	Loser		1			0.84	40%
c. /input 3/	i.	10%	Loser	1		1	1	3.04	10%
	☞ ii.	90%	Winner		1			0.84	90%
d. /input 4/	☞ i.	70%	Winner					0	70%
	ii.	30%	Loser		1			0.84	30%

The conjoined constraint $\mathbb{C}_2 \& \mathbb{C}_3$ shoulders the burden here of the superadditive gang effect: it adds a weight to the harmony score of the losing candidate in (c), making the loser's harmony score greater than the sum of the weights of $\mathbb{C}_2$ and $\mathbb{C}_3$. This allows $\mathbb{C}_2$ and $\mathbb{C}_3$ to be properly weighted below $\mathbb{C}_1$, generating the correct output candidates for (a) and (b). Furthermore, because $\mathbb{C}_1$ is now correctly weighted above $\mathbb{C}_2$ and $\mathbb{C}_3$, the correct surface output of (d) can also be modelled, where a candidate that does not violate any constraint is more likely to defeat a candidate that violates $\mathbb{C}_1$ than is a candidate that violates $\mathbb{C}_2$ or $\mathbb{C}_3$, as in (a, b).

Thus cases that will necessitate weighted constraint conjunction will be instances of extreme, superadditive gang effects in quantitative data. The following section turns to one such case, from Dioula d'Odienné, which shows cumulative effects that necessitate weighted constraint conjunction.

3 Dioula d'Odienné definiteness tone alternation

In Dioula, nouns fall into two classes of tonal behaviour for the indefinite–definite alternation.[4] In Type 1 lexical items, there is a tone change in the final vowel of the root from indefinite to definite nouns: e.g. L.L → L.H. In Type 2 lexical items, the tone change from indefinite to definite form involves both the final and penultimate syllables: e.g. L.L → H.H. I follow Braconnier (1982) in assuming that the indefinite is the underlying form from which the definite is derived. The pattern is illustrated in (10) with nouns that are underlyingly L.L in the indefinite form. (The Dioula

[4] Braconnier (1982) reports that this tone behaviour can also be triggered by a H tone in the following word, but provides no indication of the systematicity of that particular environment in conditioning the tone changes across the word boundary. Thus this paper focuses on the data available for definiteness in the lexicon. This tonally marked definite *vs*. indefinite alternation is characteristic of several Mende languages.

https://doi.org/10.1017/S0952675717000136 Published online by Cambridge University Press

data in this study comes from the lexicon in Braconnier & Diaby (1982). There are 1194 nouns in the dataset.)

(10) a. *Type 1:* L.L → L.H

indef	*def*	
fòdà	fòdá	'season'
brìsà	brìsá	'bush'
sɛ̀bɛ̀	sɛ̀bɛ́	'paper'
hàmì	hàmí	'concern'

b. *Type 2:* L.L → H.H

indef	*def*	
kùnà	kúná	'leprosy'
tùrù	túrú	'oil'
bègì	bégí	'white cotton cloth'
bìlì	bílí	'flagstone terrace'
mèlì	mélí	'worm'
sànǎ̀	sánǎ́	'tree'

In the underlyingly L-toned nouns (i.e. nouns that end with two L.L syllables), Type 1 items feature a tone change in the final syllable only: L.L → L.H. The domain of tone change in Type 2 nouns includes both the penultimate and final syllables: L.L → H.H.

In the nouns in (11) with underlying H tone, the tonal alternation is more complex, but the tone changes occur in the same domains as underlyingly L-toned nouns. Underlyingly H-toned nouns are ones that have H in the penultimate or in both the penultimate and final syllables. There are no L.H indefinite forms in Dioula.

(11) a. *Type 1:* H.H → H.LH

indef	*def*	
bɛ́sɛ́	bɛ́sɛ̌	'machete'
dáfé	dáfě	'horse'
ɟámú	ɟámŭ	'clan name'

b. *Type 2:* H.H → L.H

indef	*def*	
múrú	mùrú	'knife'
ɟégí	ɟègí	'hope'
télú	tèlú (*or* télŭ)	'tree'
ɲúmán	ɲùmán	'bond'

Underlyingly H-toned Type 1 items feature a LH contour tone on the final syllable in the definite form: H.H → H.LH. H-toned Type 2 nouns display a change across the final two syllables: H.H → L.H.

Type 1 items are ones in which only the final syllable is the domain of tone alternation in definite forms, and Type 2 items are ones in which the final two syllables are the domain of tone alternation. Following Braconnier (1982), I assume here that the definite is marked by a H-tone morpheme that must occur on at least the final syllable. The pattern is summarised in Table I.

		Type 1	Type 2
underlyingly L-toned	/L.L/	L.**H**	**H.H**
underlyingly H-toned	/H.H/	H.**LH**	**L.H**

Table I
Definiteness tone-alternation pattern in Dioula.

https://doi.org/10.1017/S0952675717000136 Published online by Cambridge University Press

3.1 Predicting Type 1 *vs.* Type 2 items

Whether an item is a Type 1 or Type 2 noun is, to an extent, predictable. In his original description of the Dioula tone system, Braconnier (1982) points to three factors that contribute to distinguishing between Type 1 and Type 2: (i) the sonority of the final intervocalic consonant, (ii) whether the place of articulation of the final two vowels is identical and (iii) the nasality or orality of the final vowel and final intervocalic consonant. Crucially, here I show quantitatively that the more similar the segments in the word-final VCV# sequence are along these independent planes of similarity, the more the noun is to be Type 2. That is, the more similarity in the final two syllables, the more likely they will be in the same tonal domain (i.e. Type 2).

Nouns with more sonorous final intervocalic consonants are more likely to be Type 2 nouns, with the domain of tone change being the final two syllables: e.g. /mèlì/ → [mélí] in (10b). This sonority effect scales quantitatively with increasing sonority (Shih 2013), as illustrated in Fig. 1. Thus, as the sonority of the final intervocalic consonant becomes more similar to the sonority of its surrounding vowels, the tonal domain for the definite alternation is more likely to include the entire VCV# sequence (i.e. the final two syllables). This is especially true when comparing the nasals ([m n ɲ ŋ]) with the liquids ([l r]) and with [g], which behaves phonotactically like a sonorant and often surfaces as a voiced fricative [ɣ] (Braconnier 1983). Braconnier (1983: 50) notes that [j] behaves phonotactically like a fricative, which corresponds to its behaviour here: nouns with [j] in final intervocalic consonant position are more likely to be Type 1. There are too few tokens of [w] (two Type 1 occurrences and one Type 2) to draw any definite conclusion about its behaviour. Obstruent consonants are the least similar in sonority to the flanking vowels, and so items with obstruent final intervocalic consonants are the least likely to be Type 2: there are only four exceptions, out of 294 total final intervocalic obstruents.

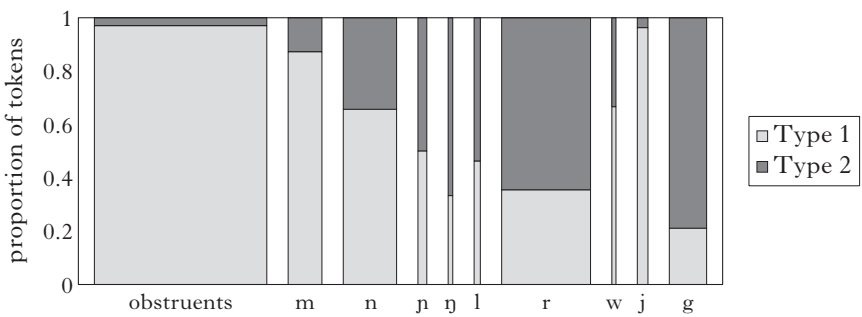


Figure 1
Final intervocalic consonants by Type.
Width of bars are scaled by proportion of data.

https://doi.org/10.1017/S0952675717000136 Published online by Cambridge University Press

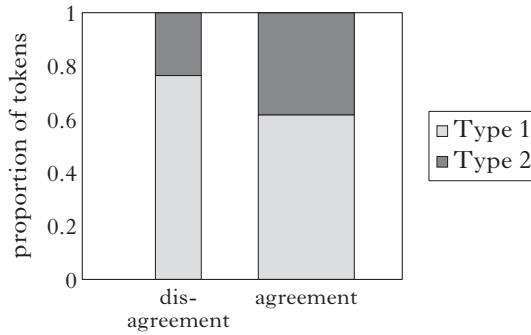


Figure 2
Nasality agreement by Type.
Width of bars are scaled by proportion of data.

Nasal similarity also contributes to distinguishing Type 1 *vs*. Type 2 items, as shown in Fig. 2. Nouns in which the final vowel and intervocalic consonant are both nasal or both oral are more likely to be Type 2 than nouns in which the final vowel and intervocalic consonant disagree in nasality or orality: /sànǎ/ → [sánǎ] in (10b).

The third predictor of Type 1 *vs*. Type 2 items noted by Braconnier (1982) is long-distance featural identity between vowels, shown in Fig. 3. If the final two vowels are featurally identical (i.e. they agree in height, rounding and backness), then the likelihood of being Type 2 increases: e.g. /bìlì/ → [bílí] in (10b).

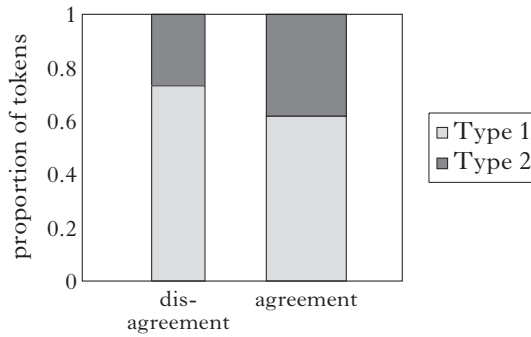


Figure 3
Vowel-identity agreement by Type.
Width of bars are scaled by proportion of data.

These similarity preconditions on definite tone behaviour gang up in a cumulative similarity interaction. The more similarity that is exhibited amongst the members of the word-final VCV# sequence in terms of

https://doi.org/10.1017/S0952675717000136 Published online by Cambridge University Press

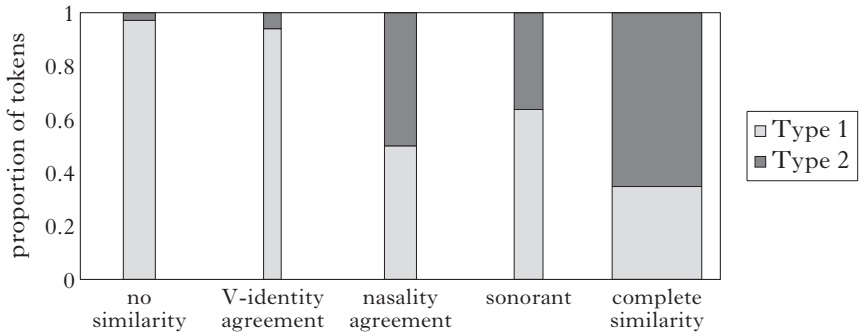


Figure 4

Final intervocalic consonants by Type.
Width of bars are scaled by proportion of data.

sonority, nasality and vowel identity, the more likely it is that the noun will be Type 2, with the tonal domain extending over the final two syllables: e.g. /tùrù/ → [túrú] in (10b). This ganging effect is illustrated in Fig. 4: the rightmost bar, which has the largest proportion of Type 2 items, represents nouns that agree in vowel identity, have a sonorant intervocalic final consonant and agree in either nasality or orality. The leftmost bar, in contrast, has the largest proportion of Type 1 items, and represents nouns that share no similarity across the VCV# sequence.[5] The bars in between represent items that only share similarity along a single dimension: from left to right, in vowel identity but not in nasality or sonority, in nasality but not in vowel identity or sonority, or in sonority but not in vowel identity or nasality.

The ganging up effect demonstrated by the Dioula nouns amounts to a superadditive phenomenon, wherein the more similar the vowels and consonant of the final two syllables are, the more likely they are to belong in the same tonal domain. The ganging effect is made up of three (somewhat) independent similarity conditions. They are independent because each contributes individual effects on Type 1 *vs*. Type 2 likelihood: for example, as is evident in Fig. 4, agreement in nasality/orality is much more likely to predict Type 2 behaviour than agreement in vowel identity alone. We will see this borne out in the constraint-weighting results in §4.

3.2 An OT approach to Dioula tone harmony

The Dioula tone pattern is formalised using Agreement by Correspondence (ABC) theory, adapted from the analysis in Shih (2013). While ABC was originally developed for long-distance consonant-harmony phenomena (Walker 2000, Hansson 2001, 2010, Rose &

[5] The dip in Type 2 representation in the graph for sonorants arises from the fact that nasal sonorants are less likely to be Type 2 than other sonorants. Lumped together, this lowers the proportion of Type 2 items for all sonorants.

https://doi.org/10.1017/S0952675717000136 Published online by Cambridge University Press

Walker 2004), it has since been extended to other segmental and tonal interactions, including vowel harmony (e.g. Sasa 2009, Walker 2009, 2014, Rhodes 2012), dissimilation (e.g. Bennett 2013, 2015a, b), local harmonies (e.g. Wayment 2009, Inkelas & Shih 2014, Lionnet 2014) and tone (e.g. Shih & Inkelas 2014, Inkelas & Shih 2016). ABC builds on the core insight that segments that are similar and proximal are more likely to interact (e.g. Kaun 1995, Burzio 2002, Zuraw 2002, Frisch *et al*. 2004). As such, ABC is particularly well suited for modelling parasitic patterns such as the similarity-based tone domains in the Dioula definiteness alternation, in which phonological similarity and proximity beget phonological interactions via correspondence domains. I focus here on the similarity bases that determine the tonal domains for the definite/indefinite alternation. For further details of the Dioula analysis that do not immediately pertain to the similarity issue, see Shih (2013).

In ABC, surface correspondence relationships are determined by the phonological similarity and proximity of segments, encoded in correspondence constraints. In essence, segments with similar features precondition the domain of phonological interaction such as agreement or dissimilation. Under an ABC approach for Dioula, the tone domains that distinguish Type 1 *vs*. Type 2 items are correspondence domains over which tone interactions take place. The set of relevant CORR constraints for Dioula nouns are given in (12). The constraints here follow pairwise correspondence and identity assessment (Hansson 2007, Rhodes 2012).[6]

(12) a. *Sonority correspondence*

CORR-X::X{V} (CORR{V})
Assign a violation for each immediately adjacent pair of segments in the output that are within the {vowel} range of the sonority scale and do not share a correspondence relationship.

CORR-X::X{V,R} (CORR{VR})
Assign a violation for each immediately adjacent pair of segments in the output that are within the {vowel, liquid} range of the sonority scale and do not share a correspondence relationship.

CORR-X::X{V,R,N} (CORR{VRN})
Assign a violation for each immediately adjacent pair of segments in the output that are within the {vowel, liquid, nasal} range of the sonority scale and do not share a correspondence relationship.

CORR-X::X (CORR)
Assign a violation for each immediately adjacent pair of segments in the output that do not share a correspondence relationship.

[6] '::' denotes immediately adjacent segments. The absence of '::' in a CORR constraint denotes segments at any distance, e.g. CORR-VV[F]. See Hansson (2001) for proximity scaling in ABC; the notation follows Inkelas & Shih (2014) for local correspondence relationships.

https://doi.org/10.1017/S0952675717000136 Published online by Cambridge University Press

b. *Nasality correspondence*
CORR-[X::X]$_{\sigma}$[±nas] (CORR[±N])
Assign a violation for each immediately adjacent pair of segments in the output that are within a syllable, are both [αnasal] and do not share a correspondence relationship.

c. *Vowel identity correspondence*
CORR-VV[F]
Assign a violation for each consecutive pair of vowels in the output that are identical in feature set [F] and do not share a correspondence relationship.

Correspondence constraints are needed for each independent dimension of similarity involved in determining the tone domain for Type 1 *vs.* Type 2 items. The constraints in (12a) mandate that segments that are sufficiently similar in sonority to adjacent segments be in a correspondence relationship; given the scalar nature of sonority, a set of scaled correspondence constraints is used here. The constraint in (12b), CORR-[X::X]$_{\sigma}$[±nas], states that segments within a syllable that share nasality or orality are in a correspondence relationship.[7] The constraint in (12c), CORR-VV[F], mandates that consecutive vowels that are featurally identical be in a correspondence relationship. For simplicity in tracking the individual similarity-based correspondence relationships, I assume here that each correspondence constraint (or set of correspondence constraints, in the case of the scalar sonority constraints) projects its own correspondence set, as demonstrated in (13) (for a similar approach, see Hansson 2014 and Walker 2014; cf. Bennett 2013 for single correspondence sets in a single word). Correspondence relationships are indicated by subscripts.[8]

(13) /t u r u/

t u_x r_x u_x	sonority	(e.g. CORR-X::X{V,R})
t u_y r u_y	vowel identity	(CORR-VV[F])
t u r_z u_z	nasality	(CORR-[X::X]$_{\sigma}$[±nas])

While correspondence constraints set up similarity-conditioned correspondence relationships, phonological interactions predicated on these correspondence sets are triggered by CORR-limiter constraints (Bennett 2013). Limiter constraints create unstable surface correspondence relationships, wherein corresponding segments are similar enough to interact but are too uncomfortably similar to stably coexist at a certain

[7] Nasality is formulated here as a bivalent feature, so that segments agreeing in *orality* also mandate similarity-induced correspondence.

[8] For readability, the same subscript is used here to indicate the overall correspondence set for each CORR; however, note that the assessment of correspondence still involves each local *pair* of segments, in accordance with Hansson (2007) and Rhodes (2012).

https://doi.org/10.1017/S0952675717000136 Published online by Cambridge University Press

distance (see Wayment 2009 and Inkelas & Shih 2014 for discussion on instability). Unstable correspondences give rise to repairs by harmony (i.e. attraction, whereby similar segments become more similar) or dissimilation (i.e. repellence, whereby similar segments become sufficiently dissimilar).

The Dioula definiteness alternation can be characterised by mandating that corresponding segments agree in tone, using the limiter constraint in (14).

(14) IDENT-XX[tone]
Assign a violation for each consecutive pair of segments that are members of a particular shared correspondence relationship that do not agree in tone specification.

Under this analysis, Type 2 behaviour in underlyingly L-toned items means that the H definite morpheme spreads onto both penultimate and final syllables: e.g. L.L → $H_x.H_x$. If the relevant CORR and limiter constraints outweigh input–output faithfulness to tone, (15), then similar segments will correspond and assimilate in tone, meaning that if the segments are similar enough, the noun will be Type 2.

(15) IDENT-IO-V[tone]
Assign a violation for each vowel that has a different tonal specification in the output than in the input.

Shih (2013) argues that underlyingly H-toned items can also follow the same correspondence-based analysis, even though the surface tonal alternations are different. The contour-tone formation in Type 1 items and lack of H-tone agreement in Type 2 items for underlyingly H-toned nouns stems from antihomophony and OCP avoidances of adjacent, heteromorphemic H tones. For simplicity here, I set aside the surface tonal differences between underlyingly H- and L-toned nouns, and focus on the similarity-based correspondence relationships, because these determine the relevant tonal domains in Type 1 *vs*. Type 2 items, which cut across any differences in underlying tones.

Hand-weighted HG tableaux are shown below as illustrations of the ABC system for Dioula tone harmony. As a simplification, the suffix H tone is assumed to be realised on the final vowel already: Shih (2013) accomplishes this in the full Dioula system using high-ranked REALISEMORPH. In (16), a sonorant final intervocalic consonant (e.g. [l]) facilitates tone agreement between the final vowel and the penultimate vowel by satisfying high-weighted CORR-X::X{V,R} and the mandated tonal identity between corresponding segments (i.e. IDENT-XX[tone]). The relevant CORR constraints that give the necessary precondition for correspondence and tone agreement must outweigh input–output faithfulness in order for the corresponding and tonally agreeing candidate, (16d), to win.

https://doi.org/10.1017/S0952675717000136 Published online by Cambridge University Press

(16)

mèlí	CORR{VR} 4	IDENT-XX[tone] 4	IDENT-IO-V[tone] 3	CORR 1	$\mathcal{H}$
a. mèlí	2			2	10
b. mè$_x$Í$_x$í$_x$		1 (VC)			4
c. mé$_x$l$_x$í$_x$		2 (VC,CV)	1		11
☞ d. mé$_x$Í$_x$í$_x$			1		3
e. mè$_x$lí$_x$	2	1		2	14
f. mé$_x$lí$_x$	2		1	2	13

In (17), a non-sonorant final intervocalic consonant t (e.g. [s]) does not facilitate tone harmony, because it is insufficiently sonorous to incur violations of CORR-X::X{V,R} when between two vocalic segments. The winner when there is an obstruent, then, is the disharmonic candidate with no correspondence in (17a).

(17)

brìsá	CORR{VR} 4	IDENT-XX[tone] 4	IDENT-IO-V[tone] 3	CORR 1	$\mathcal{H}$
☞ a. brìsá				2	2
b. brì$_x$ś$_x$á$_x$		1 (VC)			4
c. brí$_x$s$_x$á$_x$		2 (VC,CV)	1		11
d. brí$_x$ś$_x$á$_x$			1		3
e. brì$_x$sá$_x$		1		2	6
f. brí$_x$sá$_x$			1	2	5

The tableau in (18) illustrates the cumulative addition of constraint weights. In this case, sonorant consonants and featurally identical vowels facilitate tone spread, due to the additive effect of two high-weighted CORR constraints, CORR-X::X{V,R} and CORR-VV[F].

(18)

tùrú	CORR{VR} 4	CORR-VV[F] 3	IDENT-XX[tone] 4	IDENT-IO-V[tone] 3	CORR 1	$\mathcal{H}$
a. tùrú	2	1			2	13
b. tù$_{xy}$ŕ$_x$ú$_{xy}$			2 (VC,VV)			8
c. tú$_{xy}$r$_x$ú$_{xy}$			2 (VC,CV)	1		11
☞ d. tú$_{xy}$ŕ$_x$ú$_{xy}$				1		3
e. tù$_y$rú$_y$	2		1		2	14
f. tú$_y$rú$_y$	2			1	2	13

If the additive effect shown in (18) is implemented as weighted constraint conjunction on top of regular cumulative additivity (19), then there is an increase in the magnitude of harmony-score differences ($\Delta\mathcal{H}$)

https://doi.org/10.1017/S0952675717000136 Published online by Cambridge University Press

between the winner and the loser candidates. While the $\Delta\mathcal{H}$ between the harmonic winner, (d), and completely non-harmonic non-corresponding candidate, (a), in the additive approach in (18) is 10, the $\Delta\mathcal{H}$ in the super-additive approach in (19) is 14, denoting a greater predicted likelihood of tone harmony.

(19)

tùrú	Corr{VR} *Corr-VV[F] 2	Corr {VR} 4	Corr-VV[F] 3	Id-XX [tone] 4	Id-IO-V[tone] 3	Corr 1	$\mathcal{H}$
a. tùrú	2	2	1			2	17
b. $tù_{xy}ŕ_{x}ú_{xy}$				2 (VC,VV)			8
c. $tú_{xy}r_{x}ú_{xy}$				2 (VC,CV)	1		11
☞ d. $tú_{xy}ŕ_{x}ú_{xy}$					1		3
e. $tù_{y}rú_{y}$		2		1		2	14
f. $tú_{y}rú_{y}$		2			1	2	13

4 Modelling Dioula superadditivity

A question that arises is whether weighted constraint conjunctions are necessary for capturing the probabilistic similarity- and proximity-determined Dioula definite/indefinite alternation tonal domains (Type 1 = final syllable; Type 2 = final two syllables). Two candidate MaxEnt models are compared here: one without and one with weighted constraint conjunction of pairwise constraints.[9] The models were simplified by limiting output candidates to a binary choice between a Type 2 candidate that has a two-syllable correspondence domain (i.e. satisfies correspondence and all correspondence limitations) and a Type 1 candidate that has a one-syllable correspondence domain (i.e. does not satisfy correspondence). For example, a Type 2 underlyingly L-toned noun input (L.L) has two output candidates that are considered: one that maximally satisfies all correspondence constraints and agrees in tone (H.H) and one that fails to satisfy correspondence constraints, and is thus disharmonic (L.H). Because Ident-XX[tone] will be vacuously satisfied whenever Corr constraints are satisfied under this assumption, it is not shown in the results below.

Constraint-weighting results for both MaxEnt models are provided in Table II, and model statistics are given in Table III.

[9] Pairwise interactions were chosen on the basis of the evidence, as presented in §3.1 and logical combinations (e.g. it is unnecessary to have the interaction Corr-[X::X]$_{\sigma}$[±nas] × Corr-X::X{V,R,N} if the interaction Corr-[X::X]$_{\sigma}$[±nas] × Corr-X::X{V,R} already exists, because the expansion of the sonority scale to {V,R,N} would only include nasals, confounding the interaction with Corr-[X::X]$_{\sigma}$[±nas]. As observed by a reviewer, model comparison is a complex problem. The information-theoretic approach advocated for in §4.1 comes from a programme that addresses model selection using evidence- and theory-driven methods, and explicitly discourages brute-force data dredging (Anderson & Burnham 2002, Burnham & Anderson 2004).

https://doi.org/10.1017/S0952675717000136 Published online by Cambridge University Press

Without constraint conjunction		With constraint conjunction	
IDENT-IO-V[tone]	3.985	IDENT-IO-V[tone]	3.588
CORR	0.0	CORR	0.0
CORR{VRN}	1.341	CORR{VRN}	1.266
CORR{VR}	0.589	CORR{VR}	0.118
CORR{V}	2.343	CORR{V}	2.516
CORR-VV[F]	0.187	CORR-VV[F]	0.0
CORR[±N]	0.521	CORR[±N]	0.0
		CORR-VV[F] × CORR{VRN}	0.0
		CORR-VV[F] × CORR{VR}	0.263
		CORR-VV[F] × CORR[±N]	0.259
		CORR[±N] × CORR{VR}	0.466

Table II

Constraint weights: grammars with and without constraint conjunction.

	C	D_{xy}	−2 log-likelihood
without conjunction	0.8576	0.7152	1085
with conjunction	0.8654	0.7309	1059
deviance (Δ(−2 log-likelihood)): 26 (***, $df = 4$, $p < 0.0001$)			

Table III

Model statistics: grammars with and without constraint conjunction. C is the concordance statistic, which indicates the probability that predicting the outcome is better than chance. C values range between 0.5 and 1.0, with good models usually considered ones where $C > 0.8$ (Baayen 2008). D_{xy} is Somers' rank correlation, which measures the association between predicted and observed values. Measures run between 0 and 1. Algebraically lower −2 log-likelihood measures indicate higher model likelihoods.

Testing conjoined constraints reveals that the gang effect of similarity in Dioula is most active for segments that are already highly similar. For example, the conjunction CORR-VV[F] × CORR-X::X{V, R} receives a weight of 0.263, while the conjunction CORR-VV[F] × CORR-X::X{V,R,N} receives no weight: this difference indicates that the similarity effect for Type 2 items is further heightened when the VCV# segments are vowels and liquids ({V,R}) rather than vowels and nasals ({V,R,N}).

The additive effect of conjoined constraints involving liquids is further ganging. Because weighted violations of constraint conjunctions are part of the additive harmony scores, this effect amounts to superadditivity. For example, a non-agreeing VRV sequence such as *[tùrú] incurs additive

https://doi.org/10.1017/S0952675717000136 Published online by Cambridge University Press

violations not only of the simplex constraints (i.e. CORR-X::X, CORR-X::X {V,R}, CORR-X::X{V,R,N}, CORR-VV[F], CORR-[X::X]$_\sigma$[±nas]) but also the conjoined constraints (i.e. CORR-VV[F] × CORR-X::X{V,R,N}, CORR-VV[F] × CORR-X::X{V,R}, CORR-VV[F] × CORR-[X::X]$_\sigma$[±nas], CORR-[X::X]$_\sigma$[±nas] × CORR-X::X{V,R}). A model with conjunction therefore assigns harmonic [túŕú] 71% probability and disharmonic [tùrú] 29% probability, as shown in (20).

(20)

tùrú	CORR {VR} & CORR-VV[F] 0.263	CORR {VR} & CORR [±N] 0.466	CORR-VV[F] & CORR [±N] 0.259	ID-IO-V[tone] 3.588	CORR {V} 2.516	CORR {VR} 0.118	CORR {VRN} 1.266	$\mathcal{H}$	predicted
☞ a. $tú_{xy}ŕ_{x}ú_{xy}$				1				3.588	71%
b. tùrú	2	2	1			2	2	4.485	29%

In comparison, the model without conjunction in (21) only assigns harmonic [túrú] 64% and disharmonic *[tùrú] 36% probability, because it lacks the additional violations of weighted constraint conjunctions that would otherwise decrease the harmony score and predicted probability of *[tùrú].

(21)

tùrú	ID-IO-V[tone] 3.985	CORR-{V} 2.343	CORR {VR} 0.589	CORR {VRN} 1.341	CORR [±N] 0.521	CORR-VV[F] 0.187	$\mathcal{H}$	predicted
☞ a. $tú_{xy}ŕ_{x}ú_{xy}$	1						3.985	64%
b. tùrú			2	2	1	1	4.568	36%

With the conjoined constraints in (20), the model can capture the superadditive effect that goes beyond the addition of the independent constraint weights.

4.1 Comparing weighted conjunction and additive cumulativity

The necessity of weighted constraint conjunction can be tested using model comparison of grammars with and without conjoined constraints. If a gang effect is merely additive, then a grammar with conjoined constraints should not contribute any additional explanatory power. If, however, a gang effect is superadditive, then a grammar with conjoined constraints will demonstrate improved explanatory power over a grammar without conjunction.

Model-comparison tests for the necessity of constraints have been used sparingly in HG and OT approaches, largely due to the usual assumption that CON provides a constraint set and the grammar's primary concern is constraint weighting or ranking. Under this mode of operation, the

https://doi.org/10.1017/S0952675717000136 Published online by Cambridge University Press

assessment and rejection of the viability of a constraint's existence is largely left to arguments on conceptual or phonological grounds: e.g. Occam's Razor and naturalness. There are notable exceptions in the MaxEnt literature, however, that appeal to model comparison for constraint assessment: see e.g. Wilson & Obdeyn (2009), using a maximum a posteriori approach, and Hayes *et al*. (2012), using model log-likelihood comparison.

One quantitative approach to assessing competing grammars – i.e. one which allows conjunction and one which does not – is Akaike Information Criterion (AIC)-based model comparison, which allows for the comparison of significant improvements in capturing information between competing grammars. AIC model comparison is an approach founded on the idea that all models (i.e. grammars) are mere approximations of full reality, an ideal for which the true parameters (β) remain unknown (e.g. Kullback & Leibler 1951, Burnham & Anderson 2002, 2004). The aim in AIC model comparison is to reduce the amount of information loss in a candidate grammar: the less information that a candidate grammar loses, the more weight of evidence there is in favour of that particular grammar.

Information criteria measures come in various forms.[10] I use second-order AIC_c, as shown in (22), because it penalises an increasing number of constraints against a sample size, with adjustments for sample size and the number of constraints in the grammar (Burnham & Anderson 2002, 2004).

(22) $$AIC_c = -2\log(\mathcal{L}(\hat{\beta}|D)) + 2K + \frac{2K(K+1)}{n-K-1}$$

where $\mathcal{L}(\hat{\beta}|D)$ is the maximum likelihood of observed data D given fitted parameters $\hat{\beta}$
K is the number of estimable parameters (i.e. constraints) in the model
n is the sample size

When $n/K > 40$, AIC_c begins to converge with AIC. Because AIC_c regularises for sample size, it is the more conservative measure for model comparison in general (Burnham & Anderson 2004: 269–270).

Algebraically lower AIC_c values suggest better model support. As a rule of thumb in comparing candidate models, any AIC_c difference greater than 10 between two candidate models is considered large. Translated into an evidence ratio E, as in (23), a 10-point difference between two candidate models is equivalent to about a 150:1 odds that the second-best model has essentially no evidential support of being as good as the best candidate model (e.g. Anderson 2008: 89–90).

[10] Comparisons between available information criteria measures is beyond the scope of this paper. The reader is referred to the existing literature: see e.g. Burnham & Anderson (2004: 275ff) for detailed discussion of AIC *vs*. the other common IC measure, the Bayesian Information Criterion.

https://doi.org/10.1017/S0952675717000136 Published online by Cambridge University Press

(23) $E_{i,j} = \frac{1}{e^{(-(1/2)\Delta_j)}}$

for models i and j, where $\Delta_j = \mathrm{AIC}_{c_j} - \mathrm{AIC}_{c_i}$

AIC_c is based on the standard likelihood ratio test, which maximises descriptive accuracy given the observed data (see e.g. Hayes *et al*. 2012 for its use in phonology). But, unlike a likelihood ratio, AIC_c penalises a loss of restrictiveness in the grammar that potentially comes with the addition of constraint conjunction.[11] A further advantage of AIC_c comparison is that its results for assessing restrictiveness and generalisability have also been shown to asymptotically converge with k-fold cross-validation as sample size increases, while remaining computationally faster than k-fold cross-validation (e.g. Stone 1977).

It is crucial to note that the AIC_c number on its own is not a statistical test of significance or of stand-alone goodness of fit: AIC_c metrics must be taken as *comparison* statistics of more than one candidate model and between at least two AIC_c measures.

4.2 Results of model comparison

Results from AIC_c model comparison are as follows. A model with constraint conjunctions has an AIC_c value of 1101, while a model without conjunction has an AIC_c value of 1083. This 18-point difference between AIC_c scores indicates a rough 8000:1 odds (E = 8103.08) that there is substantially more support for a grammar with weighted constraint conjunctions. That is, having weighted constraint conjunctions better approximates truth in predicting Type 1 *vs*. Type 2 items in Dioula, even after penalising for increased model complexity ($\Delta K = 4$).

5 Weight of evidence for conjunction

It is not the case that AIC_c will support a more complex model in every instance. To demonstrate that AIC_c will support a simpler grammar without conjunction when there is not enough weight of evidence to justify the added parameters of conjoined constraints, I turn here to tone domains within Dioula noun stems as an illustrative case.

Indefinite forms in Dioula exhibit a similar morpheme structure constraint on tone domains to the indefinite–definite alternation tone domains discussed above. Level-tone patterns across the final two syllables of each item – i.e. H.H and L.L – tend to correlate with segments in the

[11] Wilson & Obdeyn (2009) use a maximum a posteriori approach, which in addition explicitly penalises for extreme values of estimated parameters using an assumed prior.

Also unlike likelihood ratios, AIC comparison can compare the weight of evidence for non-nested models over the same data, making it a more versatile tool in assessing models (for its use in phonology for non-nested models, see e.g. Shih 2014).

https://doi.org/10.1017/S0952675717000136 Published online by Cambridge University Press

word-final VCV sequence that are similar in sonority, vowel identity and nasality/orality. Segments that are not sufficiently similar are more likely to have a tone change at the final syllable boundary: e.g. H.L. The effect is illustrated in Fig. 5, using all disyllabic indefinite noun forms in the Dioula data ($n = 667$). The rightmost bar shows nouns that agree in vowel identity, have a sonorant final intervocalic consonant and agree in either nasality or orality. These disyllabic nouns are the ones that also demonstrate the greatest proportion of level-tone surface patterns in the indefinite form, either H.H or L.L, suggesting that segmental similarity correspondence relationships determine tone-agreement domains across syllable boundaries. The individual effects on their own in indefinites, however, do not appear to precondition a two-syllable tone domain more than having no similarity in the disyllabic sequence. As Fig. 5 shows, the similarity conditioning of tone domain is largely concentrated in the gang effect of the rightmost bar.

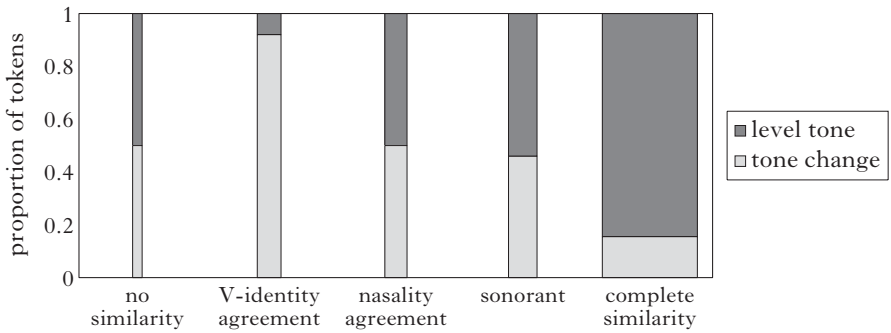


Figure 5
Cumulative similarity in disyllabic indefinite nouns.
Width of bars are scaled by proportion of data.

Using the same constraints and methodology introduced above, it is possible to examine whether constraint conjunction is needed here to achieve a superadditive effect, or whether the ganging effect observed in Fig. 5 is merely additive. Two models were tested for disyllabic indefinite items. Output candidates for the models were level *vs*. non-level tone patterns: that is, a candidate that corresponds and agrees in tone ($H_x.H_x$) *vs*. a candidate that does not satisfy correspondence ($H_x.L_y$). Because these models are predicting lexical information, no input–output faithfulness (i.e. IDENT-IO-V[tone]) was included. Resulting constraint weights are given in Table IV, with model statistics in Table V.

The high concordance statistics indicate that the models are highly accurate. A noticeable difference from the definiteness alternation results is that the weights (i.e. effect sizes) for the conjunctions are lower for this static lexical pattern. This suggests that the ganging effect here is not so extreme as to be superadditive. The AIC_c comparison results clearly corroborate this conclusion, demonstrating that a model with

https://doi.org/10.1017/S0952675717000136 Published online by Cambridge University Press

Without constraint conjunction		With constraint conjunction	
CORR	0.481	CORR	0.524
CORR{VRN}	1.054	CORR{VRN}	1.038
CORR{VR}	0.346	CORR{VR}	0.225
CORR{V}	9.536	CORR{V}	9.597
CORR-VV[F]	0.948	CORR-VV[F]	0.836
CORR[±N]	0.857	CORR-[±N]	0.699
		CORR-VV[F] × CORR{VRN}	0
		CORR-VV[F] × CORR{VR}	0.041
		CORR-VV[F] × CORR[±N]	0.232
		CORR[±N] × CORR{VR}	0.204

Table IV
Constraint weights for noun stems:
grammars with and without constraint conjunction.

	C	D_{xy}	−2 log-likelihood
without conjunction	0.983	0.967	195.95
with conjunction	0.984	0.968	195.77
deviance (Δ(−2 log-likelihood)): 0.1837 (n.s., $df = 4$, $p < 0.999$)			

Table V
Model statistics for noun stems:
grammars with and without constraint conjunction.

conjunction ($AIC_c = 220.004$) actually receives less support than a model without conjunction ($AIC_c = 212.061$). The difference in AIC_c is not large, suggesting that the effect borders on superadditive, but for disyllabic indefinite forms in the lexicon, there is insufficient weight of evidence to justify the addition of more complexity in the grammar through constraint conjunction. The support is moderate (E = 53.056; i.e. 50:1 odds) that conjunctions are not needed to describe the lexical pattern: here, we have a standard case of straightforward additive ganging that can be sufficiently captured via additive cumulativity in HG, without recourse to constraint conjunction.

6 Conclusion

This paper has demonstrated a case of superadditivity in quantitative data, where tonal domains are parasitic on the beyond-additive cumulative

https://doi.org/10.1017/S0952675717000136 Published online by Cambridge University Press

similarity of host segments. Weighted constraint conjunctions, implemented as interaction terms in Maximum Entropy Harmonic Grammar, capture superadditivity. Constraint conjunction was shown to improve the explanatory power of the grammar, even when controlling for the added complexity of conjunction.

Those who have suggested that HG additivity can supplant constraint conjunction often argue that CON should *a priori* not provide constraint conjunctions, so as to maintain restrictiveness (i.e. reduce complexity) in the constraint space (e.g. Potts *et al*. 2010, Jesney 2014). Such an argument, based on theoretical parsimony, can cut both ways, however: restrictiveness can also be maintained in the basic theoretical assumptions by allowing for an unrestricted constraint space and letting the grammar do the choosing of relevant constraints, which is arguably the grammar's task in OT. The information-theoretic model-comparison method presented here gets the best of both worlds, permitting only constraint conjunctions that are shown to improve the model, even after penalising the increased model complexity resulting from the added conjunction parameters. At the very least, it is necessary to entertain the possibility that superadditive effects are lurking in natural language data, and to quantitatively test their viability.

Weighted constraint conjunction furthermore allows a theory of conjunction a fair chance to be evaluated in a probabilistic phonological approach. Previous comparisons of conjunction and HG have been confounded by comparing only strict-ranking OT with conjunction and HG without conjunction (e.g. Potts *et al*. 2010). As shown here, once conjunction is allowed on an equal footing in a weighted grammar, it is evident that the function of constraint conjunction is not the same as mere additivity, and that the differences go beyond coreference (e.g. Pater 2016).

It is possible that embedding conjunction into HG will ultimately suggest a solution for the long-cited problem that constraint conjunction lacks an associated learnability model. With enough weight of evidence that there are cumulative effects from additive constraint interactions, a learner could consequently posit a separate and independent conjunction, thus reducing extreme values of simplex constraints. While the result would be a grammar with more complex parameters, such complexity is potentially justified by better accuracy, as measured by hypothesis (i.e. model) comparison. Thus additive cumulativity in HG can potentially guide the learning of weighted constraint conjunction for superadditive effects.

REFERENCES

Albright, Adam (2009). Cumulative violations and complexity thresholds: evidence from Lakhota. Paper presented at the 17th Manchester Phonology Meeting.

Anderson, David R. (2008). *Model based inference in the life sciences: a primer on evidence*. New York: Springer.

https://doi.org/10.1017/S0952675717000136 Published online by Cambridge University Press

Anderson, David R. & Kenneth P. Burnham (2002). Avoiding pitfalls when using information-theoretic methods. *Journal of Wildlife Management* **66**. 912–918.
Baayen, R. H. (2008). *Analyzing linguistic data: a practical introduction to statistics using R*. Cambridge: Cambridge University Press.
Baković, Eric (2000). *Harmony, dominance and control*. PhD dissertation, Rutgers University.
Bennett, William G. (2013). *Dissimilation, consonant harmony, and surface correspondence*. PhD dissertation, Rutgers University.
Bennett, William G. (2015a). *The phonology of consonants: harmony, dissimilation, and correspondence*. Cambridge: Cambridge University Press.
Bennett, William G. (2015b). Assimilation, dissimilation, and surface correspondence in Sundanese. *NLLT* **33**. 371–415.
Braconnier, C. (1982). *Le système tonal du dioula d'Odienné*. Abidjan: Université d'Abidjan.
Braconnier, C. (1983). *Phonologie du dioula d'Odienné*. Abidjan: Agence de Coopération Culturelle et Technique.
Braconnier, C. & S. Diaby (1982). *Dioula d'Odienné* (*parler de Samatiguila*)*: matérial lexical*. Abidjan: Institut de Linguistique Appliquée.
Burnham, Kenneth P. & David R. Anderson (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. 2nd edn. New York: Springer.
Burnham, Kenneth P. & David R. Anderson (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods and Research* **33**. 261–304.
Burzio, Luigi (2002). Surface-to-surface morphology: when your representations turn into constraints. In Paul Boucher (ed.) *Many morphologies*. Somerville, Mass.: Cascadilla. 142–177.
Crowhurst, Megan J. & Mark Hewitt (1997). *Boolean operations and constraint interactions in Optimality Theory*. Ms, University of North Carolina at Chapel Hill & Brandeis University. Available as ROA-229 from the Rutgers Optimality Archive.
Farris-Trimble, Ashley W. (2008). *Cumulative faithfulness effects in phonology*. PhD dissertation, Indiana University.
Frisch, Stefan A., Janet B. Pierrehumbert & Michael B. Broe (2004). Similarity avoidance and the OCP. *NLLT* **22**. 179–228.
Goldwater, Sharon & Mark Johnson (2003). Learning OT constraint rankings using a Maximum Entropy model. In Jennifer Spenader, Anders Eriksson & Östen Dahl (eds.) *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*. Stockholm: Stockholm University. 111–120.
Green, Christopher R. & Stuart Davis (2014). Superadditivity and limitations on syllable complexity in Bambara words. In Ashley W. Farris-Trimble & Jessica A. Barlow (eds.) *Perspectives on phonological theory and acquisition: papers in honor of Daniel A. Dinnsen*. Amsterdam & Philadelphia: Benjamins. 223–247.
Hansson, Gunnar Ólafur (2001). *Theoretical and typological issues in consonant harmony*. PhD dissertation, University of California, Berkeley.
Hansson, Gunnar Ólafur (2007). Blocking effects in agreement by correspondence. *LI* **38**. 395–409.
Hansson, Gunnar Ólafur (2010). *Consonant harmony: long-distance interaction in phonology*. Berkeley: University of California Press.
Hansson, Gunnar Ólafur (2014). (Dis)agreement by (non)correspondence: inspecting the foundations. *UC Berkeley Phonology Lab Annual Report: ABC Conference Archive*. Slides available (May 2017) at http://linguistics.berkeley.edu/phonlab/documents/2014/ABCC/Hansson.pdf.3-62.
Hayes, Bruce & Colin Wilson (2008). A maximum entropy model of phonotactics and phonotactic learning. *LI* **39**. 379–440.

https://doi.org/10.1017/S0952675717000136 Published online by Cambridge University Press

Hayes, Bruce, Colin Wilson & Ben George (2009). Maxent grammar tool. Software package. Available (May 2017) at http://www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool/.
Hayes, Bruce, Colin Wilson & Anne Shisko (2012). Maxent grammars for the metrics of Shakespeare and Milton. *Lg* **88**. 691–731.
Inkelas, Sharon & Stephanie S. Shih (2014). Unstable surface correspondence as the source of local conspiracies. *NELS* **44:1**. 191–204.
Inkelas, Sharon & Stephanie S. Shih (2016). Tone melodies in the age of surface correspondence. *CLS* **51**. 269–283.
Ito, Junko & Armin Mester (1998). Markedness and word structure: OCP effects in Japanese. Ms, University of California, Santa Cruz. Available as ROA-255 from the Rutgers Optimality Archive.
Ito, Junko & Armin Mester (2003). *Japanese morphophonemics: markedness and word structure*. Cambridge, Mass.: MIT Press.
Jäger, Gerhard (2007). Maximum entropy models and Stochastic Optimality Theory. In Annie Zaenen, Jane Simpson, Tracy Holloway King, Jane Grimshaw, Joan Maling & Chris Manning (eds.) *Architectures, rules, and preferences: variations on themes by Joan W. Bresnan*. Stanford: CSLI. 467–479.
Jäger, Gerhard & Anette Rosenbach (2006). The winner takes it all – almost: cumulativity in grammatical variation. *Linguistics* **44**. 937–971.
Jesney, Karen (2014). Complex process interaction: weighted constraints vs. ranked constraints and rules. Paper presented at the Interaction of Grammatical Building Blocks (IGRA): Workshop on Building Blocks, University of Leipzig. Available (May 2017) at http://www-bcf.usc.edu/~jesney/Jesney2014IGRA_handout.pdf.
Kaun, Abigail R. (1995). *The typology of rounding harmony: an optimality theoretic approach*. PhD dissertation, University of California, Los Angeles.
Kullback, S. & R. A. Leibler (1951). On information and sufficiency. *Annals of Mathematical Statistics* **22**. 79–86.
Legendre, Géraldine, Yoshiro Miyata & Paul Smolensky (1990). Harmonic Grammar: a formal multi-level connectionist theory of linguistic well-formedness: an application. In *Proceedings of the 12th Annual Conference of the Cognitive Science Society*. Hillsdale: Erlbaum. 884–891.
Lionnet, Florian (2014). Doubly triggered harmony in Laal as subphonemic Agreement by Correspondence. In John Kingston, Claire Moore-Cantwell, Joe Pater & Robert Staubs (eds.) *Proceedings of the 2013 Meeting on Phonology*. Available (May 2017) at http://journals.linguisticsociety.org/proceedings/index.php/amphonology/article/view/38.
Łubowicz, Anna (2005). Locality of conjunction. *WCCFL* **24**. 254–262.
Pater, Joe (2016). Universal Grammar with weighted constraints. In John J. McCarthy & Joe Pater (eds.) *Harmonic Grammar and Harmonic Serialism*. London: Equinox. 1–46.
Pater, Joe & Elliot Moreton (2012). Structurally biased phonology: complexity in learning and typology. *The EFL Journal* **3:2**. 1–44.
Potts, Christopher, Joe Pater, Karen Jesney, Rajesh Bhatt & Michael Becker (2010). Harmonic Grammar with linear programming: from linear systems to linguistic typology. *Phonology* **27**. 77–117.
Prince, Alan & Paul Smolensky (1993). *Optimality Theory: constraint interaction in generative grammar*. Ms, Rutgers University & University of Colorado, Boulder. Published 2004, Malden, Mass. & Oxford: Blackwell.
Rhodes, Russell (2012). Vowel harmony as Agreement by Correspondence. *Annual Report of the UC Berkeley Phonology Laboratory*. 138–168. Available (May 2017) at http://linguistics.berkeley.edu/phonlab/documents/2012/rhodes_abc_vh.pdf.
Rose, Sharon & Rachel Walker (2004). A typology of consonant agreement as correspondence. *Lg* **80**. 475–531.

https://doi.org/10.1017/S0952675717000136 Published online by Cambridge University Press

Sasa, Tomomasa (2009). *Treatment of vowel harmony in Optimality Theory*. PhD dissertation, University of Iowa.
Shih, Stephanie S. (2013). Consonant–tone interaction as Agreement by Correspondence. Ms, Stanford University & University of California, Berkeley. Available (May 2017) at http://cogsci.ucmerced.edu/shih/shih-ctoneABC-draftms_1-18-13.pdf.
Shih, Stephanie S. (2014). *Towards optimal rhythm*. PhD dissertation, Stanford University.
Shih, Stephanie S. & Sharon Inkelas (2014). A subsegmental correspondence approach to contour tone (dis)harmony patterns. In John Kingston, Claire Moore-Cantwell, Joe Pater & Robert Staubs (eds.) *Proceedings of the 2013 Meeting on Phonology*. Available (May 2017) at http://journals.linguisticsociety.org/proceedings/index.php/amphonology/article/view/22.
Smolensky, Paul (1993). Harmony, markedness, and phonological activity. Paper presented at Rutgers Optimality Workshop. Available as ROA-87 from the Rutgers Optimality Archive.
Smolensky, Paul (2006). Optimality in phonology II: harmonic completeness, local constraint conjunction, and feature domain markedness. In Paul Smolensky & Géraldine Legendre (eds.) *The harmonic mind: from neural computation to optimality-theoretic grammar*. Vol. 2: *Linguistic and philosophical implications*. Cambridge, Mass.: MIT Press. 27–160.
Smolensky, Paul & Géraldine Legendre (eds.) (2006). *The harmonic mind: from neural computation to optimality-theoretic grammar*. 2 vols. Cambridge, Mass.: MIT Press.
Stone, M. (1977). An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *Journal of the Royal Statistical Society. Series B* (*Methodological*) **39**. 44–47.
Walker, Rachel (2000). Long-distance consonantal identity effects. *WCCFL* **19**. 532–545.
Walker, Rachel (2009). Similarity-sensitive blocking and transparency in Menominee. Paper presented at the 83rd Annual Meeting of the Linguistic Society of America, San Francisco. Slides available (May 2017) at http://www-bcf.usc.edu/~rwalker/MenomineeLSAHdt.pdf.
Walker, Rachel (2014). Prominence-control and multiple triggers in vowel harmony: an ABC analysis. *UC Berkeley Phonology Lab Annual Report: ABC Conference Archive*. 202–213. Slides available (May 2017) at http://linguistics.berkeley.edu/phonlab/annual_report/documents/2014/ABCC/Walker.pdf.
Wayment, Adam (2009). *Assimilation as attraction: computing distance, similarity, and locality in phonology*. PhD dissertation, Johns Hopkins University.
Wilson, Colin (2006). Learning phonology with substantive bias: an experimental and computational study of velar palatalization. *Cognitive Science* **30**. 945–982.
Wilson, Colin & Marieke Obdeyn (2009). Simplifying subsidiary theory: statistical evidence from Arabic, Muna, Shona, and Wargamay. Ms, Johns Hopkins University.
Zuraw, Kie (2002). Aggressive reduplication. *Phonology* **19**. 395–439.

https://doi.org/10.1017/S0952675717000136 Published online by Cambridge University Press